

# EULERIAN DIGRAPHS AND DYCK WORDS, A BIJECTION

PIETRO CODARA, OTTAVIO M. D'ANTONA, MARCO GENUZIO

**ABSTRACT.** The main goal of this work is to establish a bijection between Dyck words and a family of Eulerian digraphs. We do so by providing two algorithms implementing such bijection in both directions. The connection between Dyck words and Eulerian digraphs exploits a novel combinatorial structure: a binary matrix, we call *Dyck matrix*, representing the cycles of an Eulerian digraph.

## 1. BACKGROUND, AND MOTIVATION

A digraph  $G$  is *Eulerian* if at every vertex the in-degree equals the out-degree. (Note that we do not require  $G$  to be connected.) The edge set of an Eulerian digraph  $G$  can be partitioned into directed cycles. For a non-empty multiset  $\mathbf{s} = \{s_1, s_2, \dots, s_n\}$ , of  $n$  positive integers, we call an Eulerian digraph  $\mathbf{s}$ -labelled if its edge set is partitioned into  $n$  directed cycles of length  $s_1, s_2, \dots, s_n$ , each with a distinguished first edge (and hence a unique second, third, etc.,  $m$ -th edge). Figure 1 shows a  $\{3, 2, 1\}$ -labelled Eulerian digraph, with its 3 directed cycles of size 1, 2 and 3; the  $j^{\text{th}}$  edge of the  $i^{\text{th}}$  cycle is labelled  $e_{i,j}$ . (Notice that in next Sections we endow these digraphs with a linear order on their cycles.)

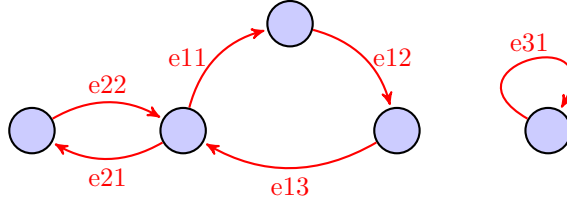
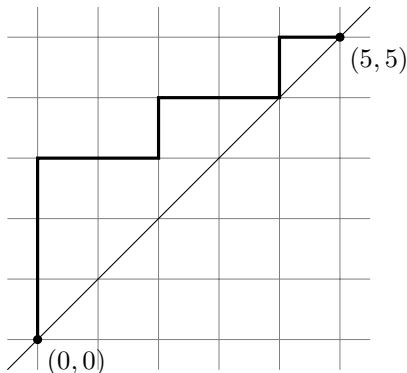


FIGURE 1. A  $\{3, 2, 1\}$ -Eulerian digraph.

A *Dyck word* on the alphabet<sup>1</sup>  $\{x, D\}$  is a string with the same number of  $x$ 's and  $D$ 's, and such that the number of  $x$ 's in any initial segment is greater or equal to the number of  $D$ 's. A *Dyck path* is a lattice path in  $\mathbb{Z}^2$  starting at  $(0, 0)$ , ending on the diagonal  $y = x$ , with unit steps in the North and East directions, and such that it does not pass below the  $y = x$  diagonal. One can easily build a correspondence between Dyck paths and Dyck words, by mapping a North step to the character  $x$  and an East step to the character  $D$ . Figure 2 shows the Dyck path associated with the word  $xxxDDxDDxD$ . We denote by  $\mathcal{W}$  the set of all Dyck words on the alphabet  $\{x, D\}$ .

<sup>1</sup>Date: July 10, 2014.

<sup>1</sup>The choice of the alphabet follows [EGH13].

FIGURE 2. The Dyck path  $xxxDDxDDxD$ .

The main goal of this work is to establish a bijection between a family of **s**-labelled Eulerian digraphs and the set  $\mathcal{W}$ . To this end, we provide two algorithms implementing such bijection in both directions. The connection between Dyck words and Eulerian digraphs exploits a novel combinatorial structure: a binary matrix, we call Dyck matrix, representing the cycles of an Eulerian digraph.

Think of the symbol  $x$  as multiplication times  $x$ , and of the symbol  $D$  as differentiation with respect to  $x$ . For each Dyck word  $w$ , the sequence of Stirling numbers of the second kind of  $w$ , written  $S_w(k)$ , can be defined as the (unique) sequence satisfying the following identity

$$(1) \quad wf(x) = \sum_{k \geq 0} S_w(k) x^k D^k f(x).$$

Observe that, for  $w = (xD)^n$ , the unique sequence satisfying (1) is the sequence  $S(n, k)$  of Stirling numbers of the second kind. Thus, the sequences  $S_w(k)$  generalize Stirling numbers of the second kind. For more details please refer to [BF12, EGH13], and references therein.

The results obtained in this work are the basis for a description of Stirling numbers of Dyck words in terms of Eulerian digraphs.

The **s**-labelled Eulerian digraphs are investigated in [CDH14], in the special case of  $n$  cycles of length  $m$ . There, the authors show that **s**-labelled Eulerian digraphs with  $k$  vertices represent a combinatorial interpretation of the generalized Stirling numbers  $S_m(n, k)$  introduced in [BPS03]. Specifically, in [CDH14, Theorem 3.1] they prove the correspondence between **s**-labelled Eulerian digraphs and colourings (hence stable partitions) of disjoint unions of cliques of the same size.

Later, Engbers, Galvin, and Hilyard ([EGH13]) pointed out that the structures investigated in [CDH14] constitute a combinatorial interpretation of the Stirling numbers of the second kind associated with Dyck words of the form  $w = (x^m D^m)^n$ .

## 2. A FAMILY OF EULERIAN DIGRAPHS, AND THEIR CYCLE MATRICES

The main data structure our algorithms are based on is a binary matrix, defined as follows.

**Definition 2.1.** A non-empty binary matrix  $M = (m_{i,j})$  of size  $n \times k$ , is a *Dyck matrix* if it satisfies the following conditions.

- (M1) There exists  $0 < h \leq k$  such that  $m_{1,j} = 1$  if and only if  $j \leq h$ .
- (M2) For each  $1 \leq i < n$ , there exist  $1 \leq a_i \leq b_i < c_i \leq k$  such that  $a_i$  is the smallest index satisfying  $m_{i,a_i} = 1$  and  $m_{i+1,a_i} = 0$ ,  $b_i$  is the greatest index such that  $m_{i,b_i} = 1$ , and  $c_i$  is the greatest index such that  $m_{i+1,c_i} = 1$ . Moreover, the following hold:
  - (M2.1)  $m_{i+1,j} = m_{i,j}$ , for  $j = 1, \dots, a_i - 1$ ;
  - (M2.2)  $m_{i+1,j} = 0$ , for  $j = a_i, \dots, b_i$ ;
  - (M2.3)  $m_{i+1,j} = 1$ , for  $j = b_i + 1, \dots, c_i$ ;
  - (M2.4)  $m_{i+1,j} = 0$ , for  $j = c_i + 1, \dots, k$ .

For example, the matrix

$$(2) \quad \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is a Dyck matrix, while the matrix

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is not, in that it violates condition (M2).

A Dyck matrix  $M = (m_{i,j})$  of size  $n \times k$  can be associated to an Eulerian digraph  $E$  with  $n$  cycles and  $k$  vertices, endowed with an order on its cycles. For  $i = 1, \dots, n$ , the  $i^{\text{th}}$  row represents the  $i^{\text{th}}$  cycle  $C_i$ . Specifically, if  $U = \{j_1, j_2, \dots, j_{s_i}\}$  is the non-empty set of indices such that  $m_{i,j} = 1$  if and only if  $j \in U$ , then  $v_{j_1}, v_{j_2}, \dots, v_{j_{s_i}}$  are the vertices of  $C_i$ , and  $(v_{j_1}, v_{j_2}), (v_{j_2}, v_{j_3}), \dots, (v_{j_{s_i-1}}, v_{j_{s_i}}), (v_{j_{s_i}}, v_{j_1})$  are its edges. If  $U$  is a singleton, then  $C_i$  is a loop. We say that  $M$  is the *cycle matrix* of  $E$ . One can easily check that the matrix represented in (2) is the cycle matrix of the Eulerian digraph depicted in Figure 1.

We can characterize the family  $\mathcal{E}_W$  of Eulerian digraph associated to Dyck matrices, as follows. Let  $\mathbf{s} = (s_1, s_2, \dots, s_n)$ , with  $s_i > 0$  for each  $i = 1, \dots, n$ . Denote by  $C_1, \dots, C_n$  the cycles of  $E$ . The  $\mathbf{s}$ -labelled Eulerian digraph  $E$  belongs to  $\mathcal{E}_W$  if and only if the following conditions hold.

- (E1) No cycle is contained into another cycle.
- (E2) If two cycles  $C_i, C_{i+1}$  share  $k$  vertices, these must be the first  $k$  vertices of both cycles.

Finally, we are ready to introduce our main results. Denote by  $\mathcal{M}_W$  the class of all Dyck matrices.

**Theorem 2.2.**  $\mathcal{W}$  and  $\mathcal{M}_W$  are in bijection.

The proof of Theorem 2.2 is provided in Section 5. Next we introduce the algorithms to associate Dyck words with cycle matrices of ordered Eulerian digraphs, hence Dyck matrices, and viceversa.

### 3. FROM DYCK WORDS TO EULERIAN DIGRAPHS

We supply an online algorithm that converts a Dyck word in the corresponding Dyck matrix. The algorithm implements an incremental construction of the result. Here, the idea is to split a Dyck word into *slopes*, *i.e.* maximal continuous sequences of  $x$ 's, and *descents*, *i.e.* maximal continuous sequences of  $D$ 's (see [EGH13]). We

call *valley* a descent followed by a slope, and *peak* a slope followed by a descent. Every peak represents a cycle. For every cycle the number of  $x$ 's from the beginning of the *slope* to its end is the number of new vertices, with respect to the previous cycle, while the difference between the number of  $D$ 's and the number of  $x$ 's represent the number of nodes shared with the next cycle. Our algorithm incrementally builds the matrix, generating a new row at the end of each descent.

The function `getMatrix`, implementing the online conversion algorithm is shown below. The function receives in input a stream of characters, the Dyck word, and outputs the associated Dyck matrix.

In line 1 we initialize the four variables  $\mathcal{X}$ ,  $\mathcal{D}$ ,  $k$ , and  $prev$ .  $\mathcal{X}$  and  $\mathcal{D}$  are counters for the number of  $x$ 's and  $D$ 's, respectively; the variable  $k$  represents the number of nodes shared by two consecutive cycles;  $prev$  is the last character read. In lines 2–10 we process the stream of characters until its end (EOS), checking that  $\mathcal{D}$  never exceeds  $\mathcal{X}$  (if  $\mathcal{X} < \mathcal{D}$  the string does not represent a Dyck word). Line 3 reads the next character of the stream: if this is a  $D$ , we set  $prev$  to “ $D$ ” and increase  $\mathcal{D}$ . Otherwise (that is, if the next character is  $x$ ), we check (line 5) whether the  $x$  follows an  $x$  or a  $D$ . In the former case, we simply increase  $\mathcal{X}$ , while in the latter we can create a new *row* and insert it in the Dyck matrix. The creation of a new row (line 7, and line 11) implements the following steps.

(R1) We create a copy  $R$  of the previous row. We modify  $R$  maintaining only the first  $k$  1's, and resetting (to 0) the others. (If we are creating the first row,  $R$  is an empty array.)

(R2) We append to  $R$  a sequence of  $\mathcal{X} - k$  1's.

The new *row* will have  $\mathcal{X}$  1's. In line 8 we insert the new *row* into the matrix  $M$ . We also need to append a sequence of  $\mathcal{X} - k$  0's to each previous row. In line 9, we update all the variables, setting  $prec$  to “ $x$ ”,  $k$  to  $\mathcal{X} - \mathcal{D}$ ,  $\mathcal{X}$  to  $k + 1$ , and  $\mathcal{D}$  to 0. The last row is created and added to the matrix in lines 11–12.

---

```

getMatrix(s)
1  $\mathcal{X} = 0$ ;  $\mathcal{D} = 0$ ;  $k = 0$ ;  $prev = "x"$ ;
2 repeat
3   if NextChar(s) is “D” then
4     |  $prev = "D"$ ;  $\mathcal{D} = \mathcal{D} + 1$ ;
   else
5     | if  $prev = "x"$  then
6       |  $\mathcal{X} = \mathcal{X} + 1$ ;
       else
7         | create a new row ;
8         | add row to  $M$  and fill previous rows with 0's;
9         |  $prec = "x"$ ;  $k = \mathcal{X} - \mathcal{D}$ ;  $\mathcal{X} = k + 1$ ;  $\mathcal{D} = 0$ ;
       end
   end
10  remove first character of  $s$  ;
    until NextChar(s)  $\neq$  EOS and  $\mathcal{X} \geq \mathcal{D}$ ;
11 create last row;
12 add last row to  $M$  and fill previous rows with 0's;
13 return  $M$ ;

```

---

## 4. FROM EULERIAN DIGRAPHS TO DYCK WORDS

Here, we supply an algorithm to convert a Dyck matrix in a Dyck word. The algorithm implements an incremental construction of the Dyck word. The main idea is to scan the whole matrix using a vertical two-value window (a  $2 \times 1$  binary matrix), to check the values of two elements lying in consecutive rows, but in the same column. For practical reasons, we add a “virtual” row of 0’s as a first row and as a last row of the matrix. After this operation we can start scanning the matrix, moving horizontally our windows till the end of each row, then skipping to the next row. Whenever the windows identifies a combination<sup>2</sup> (0, 1), we can append an  $x$  to the Dyck word we are building, because such pair represents a vertex that does not belong to any previous cycle. The pair (0, 0) does not give any information. The pair (1, 1) represents a node belonging to two consecutive cycles. The pair (1, 0) denotes a vertex that belongs to a cycle, but does not to the next cycle: whenever our windows identifies this combination we append a  $D$  to the Dyck word.

The function `getDyckWord` receives in input a Dyck matrix  $M = (m_{i,j})$ . (The algorithm can also deal with a binary stream, representing the matrix read row by row.)

In line 1 we initialize the two variables  $i$  (the counter for rows) and  $dyck$  (the Dyck word to output). (Here,  $\epsilon$  denote the empty string.) Then, after adding (line 2) two rows of 0’s in the first and last position of  $M$ , we scan the matrix (line 3–8) checking the values of the windows  $(m_{i-1,j}, m_{i,j})$ . The algorithm uses the function `checkWindow(a, b)`, which takes two binary values and returns:

- “ $D$ ”, if  $a = 1$  and  $b = 0$ ;
- “ $x$ ”, if  $a = 0$  and  $b = 1$ ;
- $\epsilon$ , otherwise.

In the code below, EOR denotes the End Of a Row, while EOM denotes the End Of the Matrix. The symbol  $*$  is used to denote the catenation of two strings.

---

```

getDyckWord( $M = (m_{i,j})$ )
1  $i = 1$ ;  $dyck = \epsilon$ ;
2 Add a row of 0’s as a first and last row of  $M$ ;
3 repeat
4    $j = 0$ ;
5   repeat
6      $dyck = dyck * \text{checkWindow}(m_{i-1,j}, m_{i,j})$ ;
7      $j = j + 1$ ;
8   until EOR;
9    $i = i + 1$ ;
until EOM;
9 return  $dyck$ ;
```

---

<sup>2</sup>For convenience, we denote our two-value matrix by its transposed.

## 5. PROOF OF THEOREM 2.2, AND FURTHER REMARKS.

To prove Theorem 2.2 we show that the algorithm described in Sections 3 and 4 are corrects, in that they associate a Dyck word with a Dyck matrix, and viceversa. Moreover, we show that both algorithms implement injective maps, and that one map is the inverse of the other.

**From Dyck words to Dyck matrices.** The algorithm writes a row of the matrix at each valley, that is, when an  $x$  is read after a  $D$ . When the first valley is met (or at the end of the word, if there are no valleys) a first row is written, with  $\mathcal{X}$  1's (the number of  $x$  read, from the beginning of the word). Such row satisfies condition (M1) in Definition 2.1, and it will do so also when, in line 8, we append to the row a number of sequences of 0's. For every subsequent valley encountered, the algorithm writes a row with the following properties.

(A) The row has  $h$  1's below the ones of the previous row, and  $h$  is smaller than the total number of 1's of the previous row. The positions of such 1's coincide with the positions of the first  $h$  1's of the previous row (lines 8 and 11, described in (R1), Section 3).

(B) The row has a sequence of  $u > 1$  adjacent 1's, starting at the index  $b + 1$ , where  $b$  is the position of the last 1 in the previous row (lines 8 and 11, described in (R2), Section 3).

(C) The row has a number of 0's at the end, since the algorithm appends a number of sequences of 0's (line 8).

From (A), (B), and (C), we easily derive that the matrix satisfies the property (M2) in Definition 2.1. By construction, the mapping of Dyck words into Dyck matrices is injective.

$$\begin{array}{ccc}
 xxxDD & xxxDDxDD & xxxDDxDDxD \\
 \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}
 \end{array}$$

FIGURE 3. From  $xxxDDxDDxD$  to its cycle matrix, step by step.

**From Dyck matrices to Dyck words.**

Let  $M = (m_{i,j})$  be a Dyck matrix of size  $n \times k$ , and let  $M'$  be the matrix obtained from  $M$  by adding a first row and a last row of 0's, according with line 2 of the algorithm in Section 4.

**Claim 1.** Every column of  $M'$  is formed by

- (i) an initial non-empty segment of 0's, followed by
- (ii) a non-empty sequence of 1's, followed by
- (iii) a non-empty sequence of 0's.

To prove Claim 1, first observe that (i) trivially follows by the fact that we have added an initial row of 0's. From condition (M2.3) in Definition 2.1 we deduce that every column must contain at least a 1. Condition (M2.1) in the same Definition say that every 1 can be followed by other 1's, and every 0 by other 0's. Summarizing,

we have that every column has an initial segment formed by a non-empty sequence of 0's followed by a non-empty sequence of 1's. The addition of a final row of 0's implies that each column has a final non-empty sequence of 0's.

It remains to show that if in a column a 1 is followed by a 0, then no other 1's appear in the subsequent positions of the column. Let  $m_{r,s}$  be the first 0 that follows a 1 in the  $s^{th}$  column of  $M$ , if such element exists (if not, our claim about  $M'$  is trivially verified for the column  $s$ ). Consider again Definition 2.1: by (M2.3), together with the fact that  $b_{r-1} < c_{r-1}$  (M2), there exists  $j > s$  such that  $m_{r,j} = 1$ . Hence, for each  $i > r$ , the condition in (M2.3) does not apply for the elements  $m_{i,s}$ . Instead, only conditions in (M2.1) or (M2.2) can apply: in both cases  $m_{i,s} = 0$ .

The algorithm in Section 4 scans  $M'$  by rows using a vertical  $2 \times 1$  window. Whenever the algorithm hits the combination (0,1), it appends an  $x$  to a word  $W$  (starting by the empty word), and whenever it hits the combination (1,0), it appends a  $D$ . By Claim 1,  $W$  starts with  $x$ , and contains exactly  $n$   $x$ 's and  $n$   $D$ 's. Moreover, at any position the number of  $D$ 's can never exceed the number of  $x$ 's. Indeed, the scan can not hit the combination (1,0) before hitting (0,1) on the same column. Hence,  $W$  is a Dyck word.

We do not prove here that the map implemented by the algorithm is injective. We show, instead, that the maps implemented by the two algorithms are one the inverse of the other.

Let  $W$  be a Dyck word with at least one valley (if there are no valley, our claim easily follows). We follow row by row the action of the function **getMatrix** on input  $W$ , and the action of **getDyckWord** on the rows that are written. The aim is to show that **getDyckWord** returns exactly the word  $W$ .

Recall that **getMatrix** writes a row of the matrix at each valley. When the first valley is met, a sequence of 1's is written in the first row (after that, a number of 0's will follow, till the end of the row). The number of these 1's equals the number  $t$  of  $x$ 's of the first slope of  $W$ . Since the reverse algorithm adds a row of 0's as a first row, when it scans the first two rows of its matrix it begins writing a Dyck word by appending  $t$   $x$ 's to the empty string.

When **getMatrix** finds a second valley (or the end of  $W$ ), it writes the second row of the matrix, according to (A), (B), and (C) in the previous paragraph. According to (A), the beginning of the row is a copy of the previous row. When scanning this part of the row, the reverse algorithm will do nothing.

According to (A) and (B), the following part of the new row is formed by a sequence of 0's, and at least one of this 0's lies below a 1. In correspondence of such 0's the reverse algorithm will append some  $D$ 's to its output. By (R1) in Section 3 the number of 0's lying below 1's is  $\mathcal{X} - k = \mathcal{D}$ . Thus, it coincides with the number of  $D$  of the first descent of  $W$ .

According to (B), a non-empty sequence of 1's is written in the following part of the row. Such 1's lie below 0's, and, by (R2), the number of such 1's equals the number of  $x$  in the second slope. Hence, when the reverse algorithm scans this part of the row, it appends to its output string the correct number of  $x$ .

The same occurs until the last row of the matrix is written. At this point the output of the reverse algorithm coincides with  $W$  without its final descent. But **getDyckWord** adds a final row of 0's to the matrix. When this final row is scanned,

`getDyckWord` appends a sequence of  $D$ 's to its output. Such  $D$ 's make the output string a correct Dyck word. Hence, the output must coincide with  $W$ .

## 6. CONCLUSION, AND FUTURE WORK

We have provided a bijection between the set of all Dyck words and a class of binary matrices, we call Dyck matrices. Dyck matrices are the cycle matrices of a family of  $\mathbf{s}$ -labelled Eulerian digraphs, endowed with an order on their cycles.

Further work aims to describe the Stirling numbers of a Dyck word (see [EGH13]) in terms of Eulerian digraphs. Indeed, it seems possible, following the same approach as in [CDH14], to obtain, for any Dyck word  $w$ , the Stirling number  $S_w$  as the collection of  $\mathbf{s}$ -labelled Eulerian digraph obtained by taking the graph  $E$  associated with  $w$  by the algorithm described in Section 3 and applying appropriate transformations on  $E$ .

## REFERENCES

- [BF12] P. Blasiak and P. Flajolet, *Combinatorial models of creation-annihilation*, Séminaire Lotharingien de Combinatoire **65** (2010/12), Art. B65c, 78.
- [BPS03] P. Blasiak, K. A. Penson, and A. I. Solomon, *The boson normal ordering problem and generalized Bell numbers*, Annals of Combinatorics **7** (2003), no. 2, 127–139.
- [CDH14] P. Codara, O. M. D'Antona, and P. Hell, *A simple combinatorial interpretation of certain generalized Bell and Stirling numbers*, Discrete Mathematics **318** (2014), no. 1, 53–57.
- [EGH13] J. Engbers, D. Galvin, and J. Hilyard, *Combinatorially interpreting generalized Stirling numbers*, arXiv:1308.2666v3 [math.CO] (2013).

(P. Codara, O. M. D'Antona, M. Genuzio) DIPARTIMENTO DI INFORMATICA, UNIVERSITÀ DEGLI STUDI DI MILANO VIA COMELICO 39/41, I-20135 MILANO, ITALY  
*E-mail address*, P. Codara, O. M. D'Antona: `{codara,dantona}@di.unimi.it`  
*E-mail address*, M. Genuzio: `marco.genuzio@studenti.unimi.it`